arXiv:2603.01143v1 [cs.CV] 1 Mar 2026

TC-SSA：基于语义槽聚合的令牌压缩用于十亿像素病理推理

预印本

Zhuo Chen*
[1]深圳先进技术大学，深圳，广东，中国[2]宁波诺丁汉大学，宁波，浙江，中国 FoSE，数学与应用数学系
smyzc9@nottingham.edu.cn

Shawn Young *
[1]深圳先进技术大学，深圳，广东，中国

Lijian Xu†深圳先进技术大学，中国广东深圳xulijian@suat-sz.edu.cn

摘要

大型视觉语言模型在计算病理学中的应用为诊断助手带来了巨大希望，但也面临着关键的计算瓶颈：全视野切片图像（WSI）的十亿像素规模。单张 WSI 通常包含超过 10^5 个图像块，导致序列长度超出标准 Transformer架构的限制。现有解决方案往往采用空间采样，但这可能会丢弃诊断关键证据。为此，我们提出 TC-SSA（通过语义槽聚合实现令牌压缩），这是一种可学习的令牌压缩框架，可将图像块特征聚合到固定数量的语义槽中。门控路由模块通过稀疏Top-2路由将图像块分配到各槽中，随后进行加权聚合，从而在严格的令牌预算下实现全局切片覆盖。生成的表示保留了与诊断相关的信息，同时将视觉令牌数量减少至原序列的 1.7%。在 SlideBench(TCGA) 上，我们的模型总体准确率达到 78.34%，在诊断子集上的准确率达到 77.14%，在令牌预算相当的情况下优于基于采样的基线方法。该方法还可泛化至多实例学习分类，在 TCGA-BRCA 上的曲线下面积达到 95.83%，在 TCGA-NSCLC 上达到 98.27%，在 PANDA 上达到 79.80%。这些结果表明，可学习的语义聚合能够在十亿像素病理推理中有效权衡效率与诊断性能。我们的代码可在 https://anonymous.4open.science/r/TC-SSA/ 获取。

关键词 十亿像素病理 · 令牌压缩 · 视觉语言模型。

1 引言

视觉语言模型（VLMs）的最新进展 Yang 等人 [2025a,b], Yu 等人 [2025], Yang 等人 [2025c], Xu 等人 [2023]，使计算病理学中的多模态推理有了新的形式 [Lu 等，2024a，Yang 等人，2025d]。这些模型能够结合视觉证据与文本查询，为切片级问答和临床辅助提供一种潜在的交互方式。然而，全切片图像（WSI）带来了一个根本性的可扩展性

*Zhuo Chen与Shawn Young对本文贡献相同，并在深圳先进技术大学实习期间完成了这项工作。

†通讯作者。

挑战。单张全切片图像包含超过 10^5 个图像块，序列长度也超出了标准 Transformer 架构的内存和计算能力限制。因此，直接对所有图像块进行端到端处理并不可行。

尽管视觉语言模型通过多样化的预训练和统一架构，近年来已发展出能力强大的病理学助手 [Huang 等人，2023，Xu 等人，2024，Chen 等人，2024，Lu 等人，2024a,b，Sun 等人，2025，Zhang 等人，2023，Yang 等人，2026, 2023], 但要将这些综合模型部署到十亿像素切片上，就必须在计算效率与诊断效果之间取得关键平衡。为应对这一挑战，现有方法主要沿着两个方向发展。首先，以 LLaVA-Med [Li 等人，2023] 和 Quilt-LLaVA [Seyfioglu 等人，2024], 为代表的空间采样策略，通过丢弃大多数图像块，将输入限制在固定的上下文窗口内，因此存在遗漏诊断关键区域的严重风险。其次，SlideChat [Chen 等人，2025a] 等稀疏注意力框架保留了更广泛的视觉证据，但推理成本显著更高。与这两个主要方向并行的是，基础性的弱监督建模通常被表述为多实例学习 [Lu 等人，2021，Shao 等，2021，Hashimoto 等人，2024，Wu 等人，2025，Yang 等人，2024], 近期研究还广泛探索了令牌压缩、指令微调和摘要技术，以支持病理模型高效开展切片级学习 [Tang 等人，2025，Guo 等人，2025，Chen 等人，2025b，Hu 等人，2025，Wang 等人，2026年，Yang 等人，2025e，Young 等人， 2025]。

为缓解效率瓶颈，我们提出TC-SSA（通过语义槽聚合进行令牌压缩），这是一种可学习的令牌预算框架，可将所有图像块特征聚合到固定的可学习语义槽中。为避免系统性地低估罕见但具有诊断关键意义的证据，语义亲和聚类目标进一步对槽利用率进行正则化。在固定令牌预算下，该瓶颈可保持全局切片覆盖，并消除密集注意力的二次计算代价。因此，我们的主要贡献如下：

1) **基于语义槽的令牌压缩：**我们提出一种用于全切片图像令牌压缩的语义驱动机制，根据共享的上下文相关性而非空间邻近性，将所有视觉令牌路由到一组固定的已学习语义槽中。该机制聚合稀疏的诊断关键证据，同时抑制冗余背景噪声，从而在严格的令牌预算下保留图像的全局上下文。

2) **语义槽的鲁棒正则化：**我们提出语义亲和聚类，以缓解槽位坍塌，并确保训练期间的路由稳定性。通过联合优化负载均衡损失、熵正则项和z损失，该框架严格强制实现均衡的槽利用率，并惩罚数值不稳定性。

3) **卓越的效率-性能权衡：** 在严格的视觉令牌预算下（仅占全切片图像的 1.7%），TC-SSA 在 SlideBench(TCGA) 上实现了 78.34% 的总体准确率，并将诊断子集的准确率提升至 77.14%。此外，它还展现出卓越的泛化能力，在 SlideBench(BCNB) 上达到 55.94%。

2 方法论

2.1 问题表述与方法概述

我们将全切片图像视为由大量图像块构成的集合。令预训练视觉编码器提取的这些图像块的视觉特征表示为输入序列 $X \in \mathbb{R}^{B\times N\times D}$。其中， N 表示图像块总数，单张千兆像素全切片图像通常包含超过 10^5 个图像块； D 表示特征维度。我们将问题定义为构建一个高效的压缩函数 $f\colon \mathbb{R}^{B\times N\times D} \to \mathbb{R}^{B\times K\times D}$。该函数将庞大的输入序列 X 投影为严格受限的有界表示 X'，其中 $K \ll N$ 是预先设定的视觉令牌预算。核心目标是在大幅减少令牌数量的同时，细致地保留原始图像的全局语义上下文，从而将密集诊断证据纳入视觉语言模型受限的上下文窗口。

为解决这一瓶颈，我们提出了一种可学习的令牌预算框架——基于语义槽聚合的令牌压缩（TC-SSA）。如图 1 所示，该框架由门控路由机制和以槽为中心的聚合模块组成。首先，轻量级门控机制为每个输入图像块在 K 个预定义语义槽上计算概率分布。为实现稀疏分配并控制计算成本，系统采用 Top-2 路由策略，使每个图像块最多分配给两个最优槽。随后，通过加权池化聚合路由后的图像块，构建紧凑的槽嵌入。每个聚合后的槽都由多层感知机单独优化，以生成最终的压缩序列 X'。随后，该序列被投影到下游视觉语言模型嵌入空间。为在训练过程中保持路由稳定性并防止槽位坍塌，该框架引入了辅助语义亲和聚类目标。

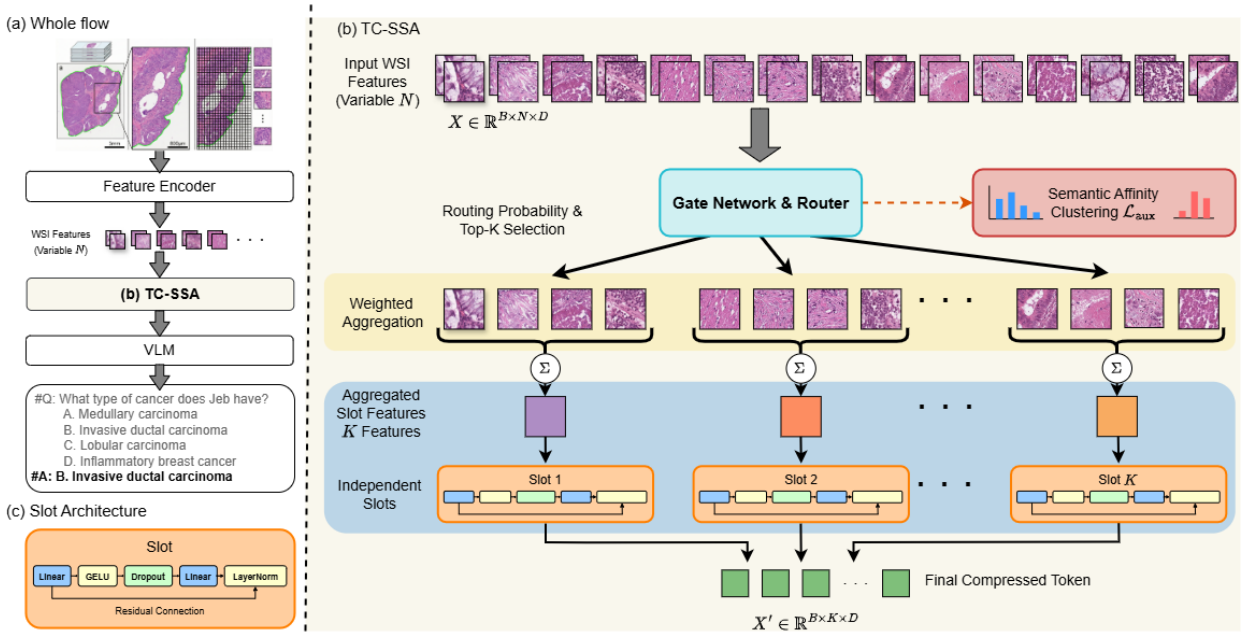


图 1：**基于语义槽聚合的令牌压缩（TC-SSA）框架概览。（a）整体流程：**特征编码器从划分后的十亿像素全切片图像块中提取可变长度的特征序列（N）。TC-SSA 模块在将这些密集特征输入视觉语言模型（VLM）进行下游诊断推理之前，将其压缩至固定的令牌预算。**（b）TC-SSA 模块：**门控网络处理输入特征 $X \in \mathbb{R}^{B\times N\times D}$，以确定路由概率，随后执行 Top-K 选择。辅助语义亲和聚类目标（$\mathcal{L}_{aux}$）对该路由分布进行正则化。经过路由的图像块特征通过加权聚合形成 K 个不同的槽，从而得到最终的压缩表示 $X' \in \mathbb{R}^{B\times K\times D}$。**（c）槽架构：**每个独立槽的内部配置。

2.2 语义槽聚合

对于输入序列 X中的每个图像块令牌 x_j，轻量级门控模块会在 K 个不同的槽上计算概率分布。该路由概率为 $P(x_j) = \text{Softmax}(W_g x_j)$，其中 $W_g \in \mathbb{R}^{B\times K\times D}$ 表示门控机制的可学习权重矩阵。为实现稀疏分配并控制计算开销，系统采用 Top-2 路由策略。该框架不会将图像块信息分配到所有可用槽，而是选出概率最高的两个槽。设 $m_{j,\,k} \in \{0,1\}$ 为一个二值掩码，用于指示第 j个图像块所选的前两个槽中是否包含第 k 个槽。因此，截断后的路由权重定义为 $P_{j,\,k} = m_{j,\,k} P_{j,\,k}$。这种硬截断确保每个图像块最多贡献给两个语义概念，从而稳定路由过程，并严格控制后续聚合成本。

在确定截断后的路由概率后，该框架执行以槽为中心的特征聚合步骤。该操作将具有相似形态或语义属性、但在空间上分散的图像块整合为统一表示。对于每个语义槽 k，通过对所有经过路由的图像块执行加权池化操作，计算得到聚合特征向量 c_k。该聚合过程表示为

$$c_k = \frac{\sum_{j=1}^{N} \tilde{P}_{j,k}\, x_j}{\sum_{j=1}^{N} \tilde{P}_{j,k} + \delta} \tag{1}$$

其中 $\delta = 10^{-9}$ 是一个较小的数值常数，用于在某个槽未接收到任何分配时防止除以零。通过对特征之和按路由权重之和进行归一化，所得令牌的幅度保持稳定，而与分配给该特定槽的图像块数量无关。该机制有效地将载玻片物理坐标空间中的高维视觉证据映射到高度压缩、固定预算的语义空间 [Yang 等人， 2026]。通过这一过程，物理上散布于组织各处的关键诊断线索被成功提炼为紧凑的视觉令牌，同时不牺牲全局上下文。

2.3 语义槽的鲁棒正则化

为防止槽位坍塌（即不成比例数量的图像块被路由到单一语义概念），该框架在训练期间引入了辅助语义亲和聚类目标。若缺乏适当的正则化，门控模块往往会严重偏向少数特定槽，从而严重限制压缩序列的表示能力。为应对这一现象，该辅助目标的主要组成部分是负载均衡损失，记为

$$\mathcal{L}_{\text{switch}} = K \sum_{k=1}^{K} P_k f_k \tag{2}$$

其中， P_k 是分配给槽 k 的平均路由概率， f_k 是在 Top-2 路由下分配给槽 k 的图像块比例。这种损失形式广泛应用于稀疏路由架构中；当分配给某个槽的图像块比例显著偏离均匀分布时，它会对网络施加惩罚。通过确保所有 K 可用槽都得到均衡利用，负载均衡损失能够保证在整个语义预算范围内充分捕捉十亿像素图像中的多样形态模式。

此外，辅助目标还整合了另外两个正则化项，以增强路由稳定性。熵正则项作用于门控机制的概率分布，以避免训练初期出现过度自信但错误的路由决策，表示如下：

$$\mathcal{L}_{\text{ent}} = 1 - \frac{-\sum_{k=1}^{K} P_k \log(P_k + \epsilon)}{\log K} \tag{3}$$

其中 $\epsilon = 10^{-8}$ 表示 nu 值稳定性。此外，z损失项 $\mathcal{L}_z$ 会惩罚过大的logit幅度 s 由门控 网络产生，从而防止数值不稳定：

$$\mathcal{L}_z = \alpha \frac{1}{N} \sum_{j=1}^{N} \left(\log \sum_{k=1}^{K} \exp(g_{j,k}) \right)^2 \tag{4}$$

其中 $g_{j,k}$ 表示 softmax 前的门控logit， $\alpha = 10^{-4}$。辅助路由正则化与下游视觉语言模型的主要任务损失（记为 $\mathcal{L}_{task}$）联合优化。具体而言，整体训练目标可表示为：

$$\mathcal{L}_{total} = \mathcal{L}_{task} + \lambda(\mathcal{L}_{\text{switch}} + 0.5\,\mathcal{L}_{\text{ent}} + \mathcal{L}_z) \tag{5}$$

其中 λ 是一个超参数，用于平衡这些辅助项的贡献。这种正则化可确保语义槽的多样性和稳定性，从而生成紧凑且信息丰富的图像块表示。

3 实验与结果

我们的模型在 SlideBench 上进行训练，遵循 [Chen 等人，2025a], ，但使用 2×A6000 GPU，而非 [Chen 等人，2025a] 所用的 8×A100。基准将**显微镜检查** 定义为低层次形态学和染色，将**诊断** 定义为用于分级和亚型分类的临床推理。**总体准确率** 是在所有多项选择视觉问答对上计算得出的。我们使用冻结的 CONCH 编码器 [Lu 等，2024a]提取图像块特征。模型使用 $K = 32$ 个槽，辅助路由损失权重为 $\lambda = 0.1$ ，并采用 Top-2 路由。我们进行了如下对比研究和消融研究。

表 1：SlideBench(TCGA) 上的性能对比，以及在 SlideBench(BCNB) 和 WSI-VQA* 上的零样本结果。SlideChatChen 等人 [2025a] 被视为完整未压缩全切片图像推理的上限。采用浮点运算次数和准确率作为指标。最佳结果以红色标出。

方法	浮点运算次数	SlideBench(TCGA)			SlideBench	WSI
		显微镜检查	诊断	总体	(BCNB)	(VQA*)
		(上界)				
SlideChat Chen 等人 [2025a]	133.3T	87.64	73.27	81.17	54.14	60.18
		（令牌压缩 ∼ 60×）				
随机基线		24.44	24.91	25.02	24.40	24.14
GPT-4o OpenAI [2024]		62.89	46.69	57.91	41.43	30.41
LLaVA-Med Li 等人 [2023]	1.70T	47.34	32.78	42.00	30.10	26.31
Quilt-LLaVA Seyfioglu 等人 [2024]	1.70T	57.76	35.96	48.07	32.19	44.43
MedDr He 等人 [2024]	1.70T	73.30	57.78	67.70	33.67	54.36
TC-SSA（我们的方法）	1.72T	**81.94**	**77.14**	**78.34**	**55.94**	**56.62**

1）效率-性能对比研究。 如表 1 所示，TC-SSA 在 SlideBench(TCGA) 上取得了最高的总体准确率，为 **78.34%** 。在令牌预算相同的情况下，TC-SSA 在诊断子集上优于 SlideChat，凸显了路由与池化策略的有效性：该策略在保留关键证据的同时去除了冗余图像块。此外，TC-SSA 的优势还进一步推广至零样本设置，在 SlideBench (BCNB) 和 WSI-VQA* 上均持续优于竞争方法，展现了我们方法的强泛化能力。

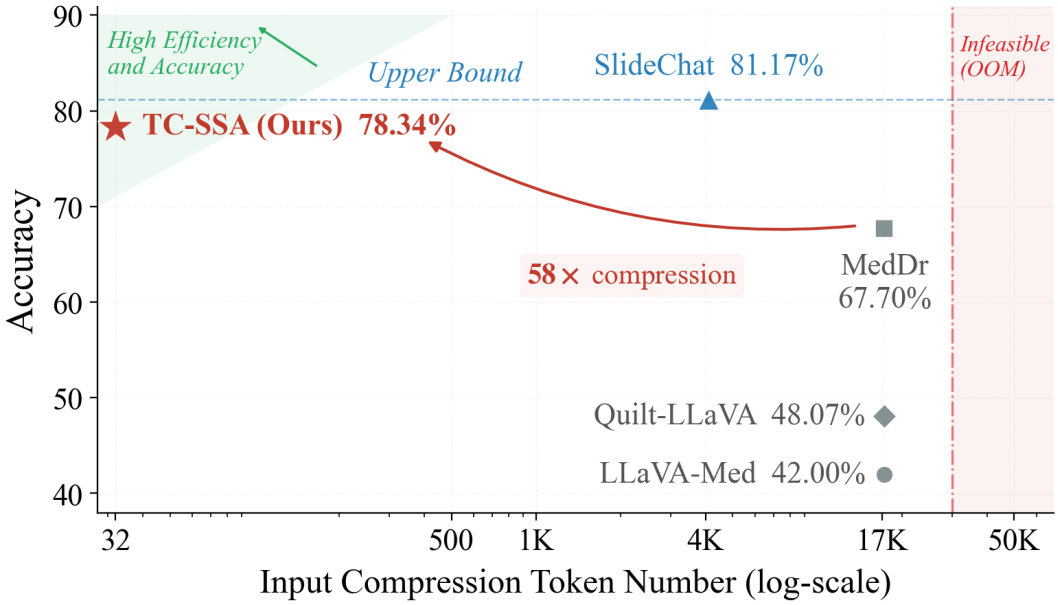


图 2：仅使用 32 个视觉令牌，TC-SSA 在 SlideBench(TCGA) 基准测试中取得了 78.34% 的总体准确率；与原始图像块特征相比，压缩比为 58× 。SlideChat 作为未经压缩的参考，而受内存不足（OOM）限制，全WSI推理无法实现。

图 2 展示了所提出的方法在 SlideBench(TCGA) 上的卓越表现。SlideChat 作为未进行显式令牌压缩时的上界，而由于 OOM 限制，全分辨率 WSI 推理并不实用。我们的方法将全局证据浓缩到 K 个令牌中，而非丢弃图像块（如 LLaVA-Med、Quilt-LLaVA），准确率比采样基线高 10.64%，且输入令牌减少 98.3%。其线性时间复杂度 $O(N \cdot K)$ 使其能够在受内存和延迟限制的临床场景中实际部署。我们进一步分析了多个TCGA队列中各槽位的语义模式。**图 3** 显示，在不同数据集中，被路由至同一槽位的图像块会在嵌入空间中聚集在一起，这表明门控机制能够持续地将相似的组织模式归为一组。

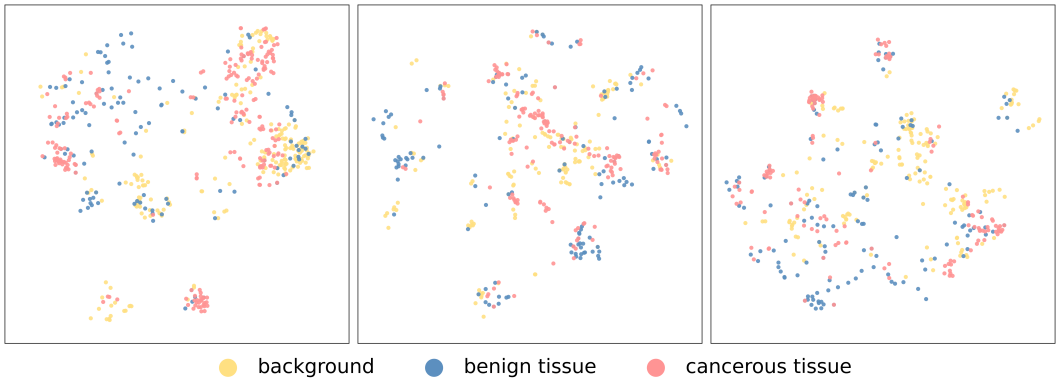


图 3：按槽位分组的多数据集图像块嵌入 t-SNE 可视化。不同颜色的簇对应不同的组织语义，表明槽位分配能够在嵌入空间中形成连贯且与数据集一致的分组。

2）消融研究。

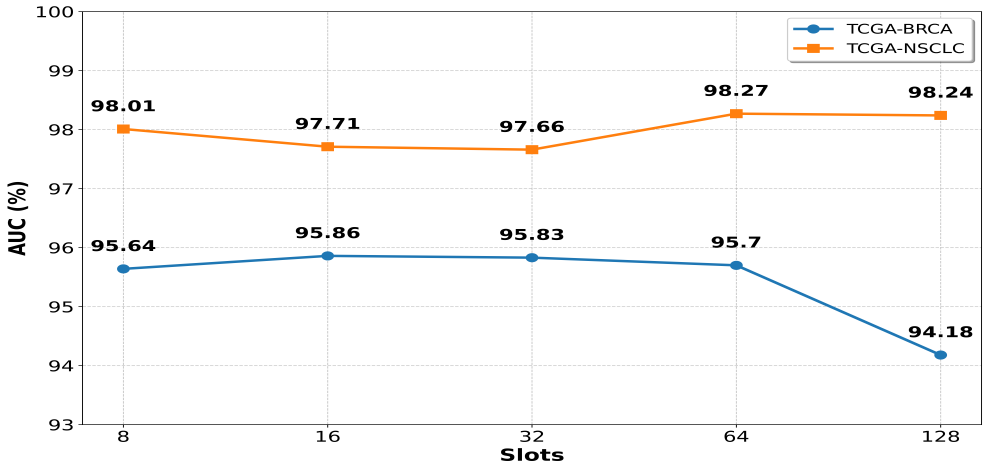


图4：在 TCGA-BRCA 和 TCGA-NSCLC 上对语义槽数量 K 进行的消融实验。我们默认设置为 $K = 32$ 。

我们首先对槽数量的选择进行了消融实验。图4展示了多实例学习曲线下面积随槽位预算 K 的变化情况。对于 TCGA-BRCA，当 $K \in \{16, 32, 64\}$ 时，性能保持稳定；当预算过大（例如 $K = 128$）时，性能则有所下降，这表明收益递减，且语义证据可能被过度碎片化。对于 TCGA-NSCLC，超过 $K = 32$ 后，增大 K 仅带来微小提升。采用 Top-2 分配并设置辅助路由损失权重 $\lambda = 0.1$ ，对于实现稳定聚类至关重要。降低 λ 会导致槽位坍塌，即单个槽吸收了超过 60% 的图像块，使模型在异质病例上的性能显著下降。增大 K 会导致槽聚合任务无关伪影和背景噪声，从而削弱传递给最终分类器的判别信号。

我们进一步开展了消融研究，将默认的 CONCH 编码器[Lu 等，2024a]替换为 UNI 编码器[Chen 等人， 2024]。值得注意的是，在采用与近期基线方法相同的 UNI 编码器时，我们的框架仍实现了最先进水平（SOTA）的性能。这表明我们的方法具有很强的泛化能力，即使使用不同于 CONCH 的编码器，也能取得极具竞争力的结果。

表 2：编码器的消融实验。在 TCGA 和 PANDA 多实例学习基准上的典型方法的曲线下面积得分。最佳结果以红色标出。

方法	TCGA-BRCA	TCGA-NSCLC	PANDA
ABMIL[Ilse 等人， 2018]	94。05%	97.04%	74.69%
TransMIL[Shao 等人， 2021]	93.33%	97.27%	68.06%
RRTMIL[Tang 等人， 2024]	94。61%	97.88%	74.93%
2DMamba[Zhang 等人， 2024]	93。08%	97。14%	76.37%
TC-SSA（本文）	**95.83%**	**98.27%**	**79.80%**

4 结论

我们提出了 TC-SSA，这是一种利用语义槽聚合的十亿像素病理视觉语言模型令牌压缩模块。在严格的视觉令牌预算下，TC-SSA 在 SlideBench(TCGA) 上达到 78.34% 的总体准确率，并将诊断子集的准确率提升至 77.14%；同时，在 TCGA-BRCA 和 TCGA-NSCLC 上分别取得 95.83% 和 98.27% 的曲线下面积。这表明，可学习聚合在受限预算下表现出色。目前，固定的切片级槽预算 K 使压缩质量取决于图块编码器，而以语义结构取代细粒度空间几何，可能会影响定位密集型任务。

参考文献

Xiaoyu Yang、Jie Lu 和 En Yu。从所有对象中学习：面向多个漂移中的多模态大语言模型自主蒸馏的概念对齐。*arXiv* 预印本 *arXiv:2510.04142*，2025a。Xiaoyu Yang、Jie Lu 和 En Yu。让多模态大语言模型从预训练阶段起适应概念漂移。载于第十三届国际学习表征会议，2025b。En Yu、Jie Lu、Kun Wang、Xiaoyu Yang 和 Guangquan Zhang。面向异构多流学习的漂移感知协同辅助混合专家。*arXiv* 预印本 *arXiv:2508.01598*，2025。Xiaoyu Yang、Jie Lu、En Yu 和 Wei Duan。非平稳漂移下的鲁棒对比预训练。*arXiv* 预印本 *arXiv:2502.07620*，2025c。Lijian Xu、Ziyu Ni、Xinglong Liu、Xiaosong Wang、Hongsheng Li 和 Shaoting Zhang。通过统一且定制的指令微调，学习用于胸部X线片解读的多任务 Transformer。*arXiv* 预印本 *arXiv:2311.01092*，2023。Ming Y Lu、Bowen Chen、Drew F K Williamson、Richard J Chen、Ivy Liang、Tong Ding、Guillaume Jaume、Igor Odintsov、Long Phi Le、Georg Gerber 等。一种用于计算病理学的视觉语言基础模型。《自然·医学》，2024a。Xiaoyu Yang、Jie Lu 和 En Yu。走钢丝：自主解耦非平稳定制微调中的有益漂移与有害漂移。载于第三十九届神经信息处理系统年会，2025d。Zhi Huang、Federico Bianchi、Mert Yuksekgonul、Thomas J Montine 和 James Zou。利用医学推特进行病理图像分析的视觉语言基础模型。《自然·医学》，2023。Hanwen Xu、Naoto Usuyama、Jaspreet Bagga、Sheng Zhang、Rajesh Rao、Tristan Naumann、Cliff Wong、Zelalem Gero、Javier González、Yu Gu 等。基于真实世界数据的数字病理全切片基础模型。《自然》，2024。Richard J Chen、Tong Ding、Ming Y Lu、Drew F K Williamson、Guillaume Jaume、Andrew H Song、Bowen Chen、Andrew Zhang、Daniel Shao 等。迈向通用计算病理学基础模型。《自然·医学》，2024。Ming Y Lu、Bowen Chen、Drew F K Williamson、Richard J Chen、Melissa Zhao、Aaron K Chow、Kenji Ikemura、Ahrong Kim、Dimitra Pouli、Ankush Patel 等。面向人体病理学的多模态生成式人工智能助手。《自然》，2024b。Yuxuan Sun、Yixuan Si、Chenglu Zhu、Xuan Gong、Kai Zhang、Pingyi Chen、Ye Zhang、Zhongyi Shui、Tao Lin 和 Lin Yang。Cpath-omni：用于计算病理学图像块与全切片图像分析的统一多模态基础模型。载于计算机视觉与模式识别会议，2025。Yunkun Zhang、Jin Gao、Mu Zhou、Xiaosong Wang、Yu Qiao、Shaoting Zhang 和 Dequan Wang。用于病理图像分类的文本引导基础模型适配。载于医学图像计算与计算机辅助干预会议，2023。Xiaoyu Yang、Lijian Xu、Xingyu Zeng、Xiaosong Wang、Hongsheng Li 和 Shaoting Zhang。Scalar：严苛开放世界中的鲁棒视觉空间概念对齐。《模式识别》，第 113203 页，2026。Xiaoyu Yang、Yufei Chen、Xiaodong Yue、Shaoxun Xu 和 Chao Ma。用于类别不平衡分类的t分布球面特征表示。载于人工智能促进协会年会论文集，第 37 卷，第 10825–10833 页，2023。Chunyuan Li、Cliff Wong、Sheng Zhang、Naoto Usuyama、Haotian Liu、Jianwei Yang、Tristan Naumann、Hoifung Poon 和 Jianfeng Gao。LLaVA-Med：在一天内训练面向生物医学的大型语言与视觉助手。载于神经信息处理系统大会，2023。Mehmet Saygin Seyfioglu、Wisdom Oluchi Ikezogwo、Fatemeh Ghezloo、Ranjay Krishna 和 Linda G. Shapiro。Quilt-LLaVA：从开源组织病理学视频中提取局部叙事，实现视觉指令微调。载于计算机视觉与模式识别会议，2024。Ying Chen、Guoan Wang、Yuanfeng Ji、Yanjun Li、Jin Ye、Tianbin Li、Ming Hu、Rongshan Yu、Yu Qiao 和 Junjun He。SlideChat：用于全切片病理图像理解的大型视觉语言助手。载于计算机视觉与模式识别会议，2025a。Ming Y Lu、Drew F K Williamson、Tiffany Y Chen、Richard J Chen、Matteo Barbieri 和 Faisal Mahmood。全切片图像上高效、弱监督的计算病理学。《自然·生物医学工程》，2021。Zhuchen Shao、Hao Bian、Yang Chen、Yifeng Wang、Jian Zhang、Xiangyang Ji 和 Yongbing Zhang。TransMIL：用于全切片图像分类的基于Transformer的相关多实例学习。载于神经信息处理系统大会，2021。

Noriaki Hashimoto、Hiroyuki Hanada、Hiroaki Miyoshi、Miharu Nagaishi、Kensaku Sato、Hidekata Hontani、Koichi Ohshima 和 Ichiro Takeuchi。利用全切片图像和流式细胞术进行淋巴瘤多实例学习分类的多模态门控专家混合。《病理学信息学杂志》，2024年。Junxian Wu、Minheng Chen、Xinyi Ke、Tianwang Xun、Xiaoming Jiang、Hongyu Zhou、Lizhi Shao 和 Youyong Kong。利用专家混合学习十亿像素全切片图像中的异质组织。载于计算机视觉与模式识别会议，2025年。Xiaoyu Yang、Lijian Xu、Simon Yu、Qing Xia、Hongsheng Li 和 Shaoting Zhang。利用基于几何的级联神经网络进行冠状动脉分割与血管矢量化。《*IEEE*医学影像学汇刊》，44(1):259–269，2024年。Wenhao Tang、Rong Qin、Heng Fang、Fengtao Zhou、Hao Chen、Xiang Li 和 Ming-Ming Cheng。重新审视计算病理学中采用切片级监督的端到端学习。载于神经信息处理系统大会，2025年。Zhengrui Guo、Conghao Xiong、Jiabo Ma、Qichen Sun、Lishuang Feng、Jinzhuo Wang 和 Hao Chen。Focus：用于少样本全切片图像分类的知识增强型自适应视觉压缩。载于计算机视觉与模式识别会议，2025年。Kaitao Chen、Mianxin Liu、Fang Yan、Lei Ma、Xiaoming Shi、Lilong Wang、Xiaosong Wang、Lifeng Zhu、Zhe Wang、Mu Zhou 等。面向病理学视觉与语言分析的高性价比指令学习。《自然·计算科学》，2025 b。Qingqiao Hu、Weimin Lyu、Meilong Xu、Kehan Qi、Xiaoling Hu、Saumya Gupta、Jiawei Zhou 和 Chao Chen。Loc-path：为病理学多模态大语言模型学习压缩。*arXiv*预印本 *arXiv:2512.05391*，2025年。Baizhi Wang、Kun Zhang、Yuhao Wang、Yunjie Gu、Haijing Luan、Ying Zhou、Taiyuan Hu、Rundong Wang、Zhidong Yang、Zihang Jiang 等。Wsisum：通过双层语义重建实现全切片图像摘要。《医学图像分析》，2026年。Xiaoyu Yang、Lijian Xu、Hongsheng Li 和 Shaoting Zhang。一叶知秋：采用语义感知视图的基于遮挡的对比学习，实现高效视觉表征。载于国际机器学习会议，第71425–71440页。机器学习研究论文集，2025e。Shawn Young、Xingyu Zeng 和 Lijian Xu。更少的令牌，更大的扩展：用于高效且可扩展表征学习的自适应视觉基底。*arXiv*预印本 *arXiv:2511.19515*，2025年。OpenAI。GPT-4o系统卡。*arXiv*预印本 *arXiv:2410.21276*，2024年。Sunan He、Yuxiang Nie、Zhixuan Chen、Zhiyuan Cai、Hongmei Wang、Shu Yang 和 Hao Chen。MedDr：面向大规模医学视觉语言学习的诊断引导式自举。*CoRR*，2024年。Maximilian Ilse、J. Tomczak 和 M. Welling。基于注意力的深度多实例学习。载于国际机器学习会议，2018年。Wenhao Tang、Fengtao Zhou、Shengyue Huang、Xiang Zhu、Yi Zhang 和 Bo Liu。特征再嵌入：迈向计算病理学的基础模型级性能。载于计算机视觉与模式识别会议，2024年。Jingwei Zhang、Anh Tien Nguyen、Xi Han、Vincent Quoc-Huy Trinh、Hong Qin、Dimitris Samaras 和 Mahdi S. Hosseini。2dmamba：用于图像表征的高效状态空间模型及其在十亿像素全切片图像分类中的应用。载于计算机视觉与模式识别会议，2024年。